

# Recommender system of physical and mathematical documents classification

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

---

## Abstract

© 2018 CEUR-WS. All rights reserved. The problems of increasing the value of scientific classifiers for the systematization of scientific information in the digital age are discussed, for example, the classification of documents (assignment of indices-classifiers) is a traditional way of systematization of knowledge and information search. In this paper we propose a recommendation system for automated selection of Universal decimal classification (UDC) indices for physical and mathematical documents. This system implements one of the services of the digital mathematical library Lobachevskii-DML. The proposed algorithm is based on the use of terms extracted from the title, the list of keywords and annotations given in the analyzed documents. Extraction of terms from the documents of the collection is carried out with the help of software tools developed by us, taking into account the stylistic features of the documents and the positions in the text of the required terms. The data obtained were included in a dictionary that has an inverted index structure. The generated dictionary contains both classification features and sets of key terms, which are used to systematize and classify the material. Most of these terms were obtained by automated processing of the collection of archives of physical and mathematical publications of the All-Russian mathematical portal Math-Net.Ru. The proposed variant of the semantic markup of a table of indices of the Universal decimal classification. The model of classification of scientific documents is described.

---

## Keywords

Classification system, Information extraction, Recommender system, Scientific information classifiers, Systematization of scientific information, UDC

## References

- [1] Bibliotechno-bibliograficheskaja klassifikacija. – URL: <http://roslavl.library67.ru/files/382/bbk.pdf>.
- [2] Gosudarstvennyj rubrikator nauchno-tehnicheskij informacii. – URL: <http://grnti.ru/>.
- [3] 2010 Mathematics Subject Classification. – URL: <http://www.ams.org/mathscinet/msc/msc2010.html>.
- [4] UDC Summary Linked Data. – URL: <http://udcdata.info/>.
- [5] Elizarov A.M., Zuev D.S., and Lipachev E.K. Upravlenie zhiznennym ciklom jelektronnyh publikacij v informacionnoj sisteme nauchnogo zhurnala // Vestnik Voronezh. gos. un-ta. Ser. Sistem. analiz i inform. tehnologij, 2014. – No. 4. – P. 81–88.
- [6] Ricci F., Rokach L., Shapira B., and Kantor P.B. (Eds.) Recommender Systems Handbook. Springer-Verlag New York. 2011. 842 p. <https://doi.org/10.1007/978-0-387-85820-3>.
- [7] Ricci F., Rokach L., and Shapira B. (Eds.) Recommender Systems Handbook. – Springer-Verlag New York, 2015. 1003 p. <https://doi.org/10.1007/978-14899-7637-6>.

- [8] Elizarov A.M., Lipachev E.K., Nevzorova O.A., and Solov'ev V.D. Methods and means for semantic structuring of electronic mathematical documents // *Doklady Mathematics*. – 2014. – Vol. 90 (1). – P. 521-524. <https://doi.org/10.1134/S1064562414050275>.
- [9] Elizarov A.M., Kirillovich A.V., Lipachev E.K., Zhizhchenko A.B., and Zhil'tsov N.G. Mathematical Knowledge Ontologies and Recommender Systems for Collections of Documents in Physics and Mathematics // *Doklady Mathematics*. – 2016. – Vol. 93 (2). – P. 231-233. <https://doi.org/10.1134/S1064562416020174>.
- [10] Elizarov A.M., and Lipachev E.K. Semanticheskie metody i instrumenty jelektronnoj matematicheskoy biblioteki Lobachevskii-DML // *Nauchnyj servis v seti Internet: trudy XIX Vserossijskoj nauchnoj konferencii (18-23 September 2017, g. Novorossiysk)*. – M.: IPM im. M.V. Keldysha, 2017. – S. 130-136. – URL: <http://keldysh.ru/abrau/2017/73.pdf>. <https://doi.org/10.20948/abrau-2017-73>.
- [11] All-Russian Mathematical Portal Math-Net.Ru. – URL: <http://www.mathnet.ru/>.
- [12] Khaydarov S.M., and Yamalutdinova G.S. Algorithm for forming the dictionary of the recommender system of selecting classifiers of scientific information // *Uchenye zapiski ISGZ*. – 2017. – Vol. 15 (1). – P. 552-557.
- [13] Ingersoll G.S., Morton T.S., and Farris D. *Taming Text: How to Find, Organize, and Manipulate It* – M.: DMK Press, 2015. – 414 p.
- [14] Manning C.D., Raghavan P., and Schütze H. *Introduction to Information Retrieval* – M.: ООО «И. Д. Вильямс», 2014. – 528 p.
- [15] Nedil'chenko O.S. Jetapy i metody avtomaticheskogo izvlechenija ključevyh slov // *Molodoj uchenyj*. – 2017. – No. 22 (156). – P. 60-62.
- [16] Reznikov I.A. Obzor algoritmov izvlechenija ključevyh slov iz teksta // 58-ja nauchnaja konferencija Moskovskogo fiziko-tehnicheskogo instituta (23-28 November 2015, Dolgoprudny). – URL: [http://conf58.mipt.ru/static/reports\\_pdf/499.pdf](http://conf58.mipt.ru/static/reports_pdf/499.pdf).
- [17] Nokel' M.A., and Lukashevich N.V. Tematicheskie modeli v izvlechenii od-noslovnih terminov // *Programmnaja inzhenerija*. – Izdatel'stvo «Novye tehnologii» (Moscow), 2015. – No. 3. – P. 34-40.